

STANISŁAW CZAJKA  
ZYGMUNT KACZMAREK

Katedra Metod Matematycznych i Statystycznych Akademii Rolniczej, Poznań  
Instytut Genetyki Roślin PAN, Poznań

## O testowaniu jednorodności współczynników zmienności\*

### On testing the homogeneity of coefficients of variation

Omawiany jest przybliżony test dla badania jednorodności współczynników zmienności, podany przez Bennetta (1976). Do testowania hipotezy zerowej o równości  $k$  współczynników zmienności proponowana jest statystyka  $Z$  posiadająca przybliżony rozkład  $\chi^2$  z  $(k-1)$  stopniami swobody. Sposób testowania równości współczynników zmienności uzupełniony jest o informacje dotyczące przydatności testu w różnych sytuacjach praktycznych, uzyskane na podstawie przeprowadzonych badań symulacyjnych.

**Słowa kluczowe:** obszar krytyczny, statystyka  $Z$ , wariancja, współczynnik zmienności

An approximate test for homogeneity of coefficients of variation given by Bennett (1976) is described. The approximate  $Z$  statistic for testing the null hypothesis of homogeneity for  $k$  coefficients of variation is proposed. It is approximately distributed as  $\chi^2$  with  $(k-1)$  degrees of freedom. Additional information concerning usefulness of the  $Z$  test in different practical situations, obtained on the basis of simulation study, is also given.

**Key words:** coefficient of variation, critical region, variance,  $Z$  statistic

### WSTĘP

W wielu problemach genetyki ilościowej dotyczących charakterystyki badanych populacji, obok miar położenia takich jak średnia arytmetyczna, mediana czy wartość modalna, wykorzystywane są miary rozproszenia (dyspersji), spośród których odchylenie standardowe, wariancja i współczynnik zmienności należą do najczęściej stosowanych. Powszechnie praktykowane jest podawanie średnich arytmetycznych i odchylenia standardowego lub wariancji dla opisu populacji o rozkładzie normalnym. Często jednak oprócz porównywania średnich dla tych populacji konieczne jest porównanie wariancji. Nie stwarza żadnego problemu porównanie zmienności dwu populacji pod względem tej samej cechy. Stosowany jest wówczas test  $F$  będący ilorazem ocen wariancji wyznaczo-

\* Praca wykonana częściowo w ramach grantu KBN 5 PO6A 01019

nych na podstawie obserwacji w próbach opisujących badane populacje. Możliwe jest także porównanie zmienności większej liczby populacji, których obserwacje dotyczą jednej cechy. Pomocny może okazać się w tym wypadku test Bartletta weryfikujący hipotezę o równości wariancji  $k$  populacji normalnych. Szczegółowy opis tego testu można znaleźć między innymi w podręcznikach Elandt (1964) i Grenia (1987). Jeśli jednak celem byłoby określenie i porównanie zmienności populacji pod względem różnych cech, najlepszą miarą wydaje się być współczynnik zmienności wyrażany zazwyczaj w procentach. Powstaje pytanie, czy tak jak w przypadku wariancji, możliwe jest testowanie równości współczynników zmienności. Przybliżony test dla jednorodności współczynników zmienności został zaproponowany przez Bennetta (1976). W pracy omówiony jest sposób testowania równości współczynników zmienności, a także przeprowadzone jest, poprzez obliczenia symulacyjne, badanie wpływu liczby obserwacji oraz liczby porównywanych współczynników zmienności na wynik testowania.

#### APROKSYMACYJNY TEST DLA JEDNORODNOŚCI WSPÓŁCZYNNIKÓW ZMIENNOŚCI

Omówimy poniżej sposób testowania jednorodności współczynników zmienności zaproponowany przez Bennetta (1976).

Weźmy pod uwagę  $k$  prób niezależnych o liczebnościach  $n_1, n_2, \dots, n_k$  pobranych z  $k$  populacji normalnych  $N(\mu_i, \sigma_i^2)$ .

Oznaczmy przez  $CV_i = \sigma_i/\mu_i$ , ( $\mu_i > 0$ ), współczynnik zmienności dla  $i$ -tej populacji ( $i = 1, 2, \dots, k$ ). Wówczas testowaną hipotezę zerową  $H_0$  o jednorodności współczynników zmienności możemy zapisać następująco:

$$H_0 : CV_1 = CV_2 = \dots = CV_k = CV = \text{const.}$$

Hipotezę alternatywną  $H_1$  jest brak równości współczynników zmienności dla co najmniej dwóch populacji. W celu przeprowadzenia weryfikacji hipotezy  $H_0$  wprowadzamy następujące oznaczenia:

przez  $x_{11}, x_{12}, x_{13}, \dots, x_{1n_1}$  — określimy zbiór obserwacji z próby dla populacji  $X_1$ , przez

$x_{21}, x_{22}, x_{23}, \dots, x_{2n_2}$  — zbiór obserwacji z próby dla populacji  $X_2$ , wreszcie przez:

$x_{k1}, x_{k2}, x_{k3}, \dots, x_{kn_k}$  — zbiór obserwacji z próby dla populacji  $X_k$ .

Niech ponadto  $n = n_1 + n_2 + \dots + n_k$  oznacza sumę wszystkich obserwacji, wielkość

$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$  — średnią z próby dla  $i$ -tej populacji,

natomiast  $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$  — wariancję z próby dla  $i$ -tej populacji.

Wówczas:

$z_i = \frac{S_i}{\bar{x}_i}$  — jest oceną współczynnika zmienności z próby dla  $i$ -tej populacji.

Odpowiednią statystyką dla testowania hipotezy  $H_0$  jest statystyka  $Z$  postaci:

$$Z = (n - k) \log \left\{ \frac{\sum_{i=1}^k \frac{n_i z_i^2}{1 + z_i^2}}{n - k} \right\} - \sum_{i=1}^k (n_i - 1) \log \left\{ \frac{\frac{n_i z_i^2}{1 + z_i^2}}{n_i - 1} \right\},$$

mająca przybliżony rozkład  $\chi^2$  z  $(k-1)$  stopniami swobody.

Hipotezę  $H_0$  o jednorodności współczynników zmienności odrzucamy na poziomie istotności  $\alpha$ , jeśli:

$$Z > \chi_{\alpha; k-1}^2.$$

### WYNIKI BADAŃ SYMULACYJNYCH

W celu sprawdzenia przydatności testu  $Z$  w różnych sytuacjach praktycznych oraz określenia wpływu liczby porównywanych populacji  $k$  i liczebności obserwacji  $n_i$  na jednorodność współczynników zmienności odpowiadających wspomnianym populacjom, wykonano badania symulacyjne. Testowanie hipotezy  $H_0$  przeprowadzono na poziomach istotności  $\alpha = 0,05$ ,  $\alpha = 0,01$  i  $\alpha = 0,001$  dla wszystkich możliwych kombinacji  $k = 2, 3, \dots, 20$  oraz  $n_i = 10, 20, \dots, 100$  przy stałej różnicy  $d$  (%) między poszczególnymi współczynnikami zmienności wynoszącej 1%, 2%, 3%, 4%, 5%, 10%, 15% i 20%. Wartości statystyki  $Z$ , wyliczone były dla każdego  $k$  i każdego  $n_i$  dotyczyły zbiorów  $CV_i$  rozpoczynających się od współczynnika zmienności równego 1% z krokiem  $r = 5\%$ .

Poniżej podajemy wybrane przykłady obliczeń symulacyjnych.

#### Przykład 1

Testowanie na poziomie istotności  $\alpha = 0,01$  jednorodności dwóch współczynników zmienności ( $H_0 : CV_1 = CV_2$ ) różniących się o 5% przy liczebnościach prób wynoszących 70. Mamy zatem:

$k = 2$ ;  $n_1 = n_2 = 70$ ;  $d = 5\%$ ;  $r = 5\%$ ;  $\alpha = 0,01$ .

Krok 1.  $CV_1 = 1\%$ ,  $CV_2 = 6\%$ ,  $Z = 67,386^{**}$  —  $H_0$  odrzucona

Krok 2.  $CV_1 = 6\%$ ,  $CV_2 = 11\%$ ,  $Z = 10,258^{**}$  —  $H_0$  odrzucona

Krok 3.  $CV_1 = 11\%$ ,  $CV_2 = 16\%$ ,  $Z = 3,971$  —  $H_0$  nieodrzucona

$\chi_{0,01; 1}^2 = 6,635$ .

#### Przykład 2

Testowanie na poziomie istotności  $\alpha = 0,01$  jednorodności trzech współczynników zmienności ( $H_0 : CV_1 = CV_2 = CV_3$ ) różniących się o 10% przy liczebnościach prób wynoszących 90. Mamy zatem:

$k = 3$ ;  $n_1 = n_2 = n_3 = 90$ ;  $d = 10\%$ ;  $r = 5\%$ ;  $\alpha = 0,01$

Krok 1.  $CV_1 = 1\%$ ,  $CV_2 = 11\%$ ,  $CV_3 = 21\%$ ,  $Z = 184,198^{**}$  —  $H_0$  odrzucona

Krok 2.  $CV_1 = 6\%$ ,  $CV_2 = 16\%$ ,  $CV_3 = 26\%$ ,  $Z = 62,708^{**}$  —  $H_0$  odrzucona

Krok 3.  $CV_1 = 11\%$ ,  $CV_2 = 21\%$ ,  $CV_3 = 31\%$ ,  $Z = 33,655^{**}$  —  $H_0$  odrzucona  
 Krok 4.  $CV_1 = 16\%$ ,  $CV_2 = 26\%$ ,  $CV_3 = 36\%$ ,  $Z = 20,712^{**}$  —  $H_0$  odrzucona  
 Krok 5.  $CV_1 = 21\%$ ,  $CV_2 = 31\%$ ,  $CV_3 = 41\%$ ,  $Z = 13,740^{**}$  —  $H_0$  odrzucona  
 Krok 6.  $CV_1 = 26\%$ ,  $CV_2 = 36\%$ ,  $CV_3 = 46\%$ ,  $Z = 9,564^{**}$  —  $H_0$  odrzucona  
 Krok 7.  $CV_1 = 31\%$ ,  $CV_2 = 41\%$ ,  $CV_3 = 51\%$ ,  $Z = 6,884$  —  $H_0$  nieodrzucona  
 $\chi^2_{0,01; 2} = 9,210$ .

### Przykład 3

Testowanie na poziomie istotności  $\alpha = 0,05$ , jednorodności trzech współczynników zmienności ( $H_0 : CV_1 = CV_2 = CV_3$ ) różniących się o 20% przy liczebnościach prób wynoszących 30. Mamy więc:

$k = 3$ ;  $n_1 = n_2 = n_3 = 30$ ;  $d = 20\%$ ;  $r = 5\%$ ;  $\alpha = 0,05$ .

Krok 1.  $CV_1 = 1\%$ ,  $CV_2 = 21\%$ ,  $CV_3 = 41\%$ ,  $Z = 75,260^*$  —  $H_0$  odrzucona  
 Krok 2.  $CV_1 = 6\%$ ,  $CV_2 = 26\%$ ,  $CV_3 = 46\%$ ,  $Z = 32,458^*$  —  $H_0$  odrzucona  
 Krok 3.  $CV_1 = 11\%$ ,  $CV_2 = 31\%$ ,  $CV_3 = 51\%$ ,  $Z = 20,033^*$  —  $H_0$  odrzucona  
 Krok 4.  $CV_1 = 16\%$ ,  $CV_2 = 36\%$ ,  $CV_3 = 56\%$ ,  $Z = 13,564^*$  —  $H_0$  odrzucona  
 Krok 5.  $CV_1 = 21\%$ ,  $CV_2 = 41\%$ ,  $CV_3 = 61\%$ ,  $Z = 9,636^*$  —  $H_0$  odrzucona  
 Krok 6.  $CV_1 = 26\%$ ,  $CV_2 = 46\%$ ,  $CV_3 = 66\%$ ,  $Z = 7,059^*$  —  $H_0$  odrzucona  
 Krok 7.  $CV_1 = 31\%$ ,  $CV_2 = 51\%$ ,  $CV_3 = 71\%$ ,  $Z = 5,285$  —  $H_0$  nieodrzucona

$\chi^2_{0,05; 2} = 5,991$ .

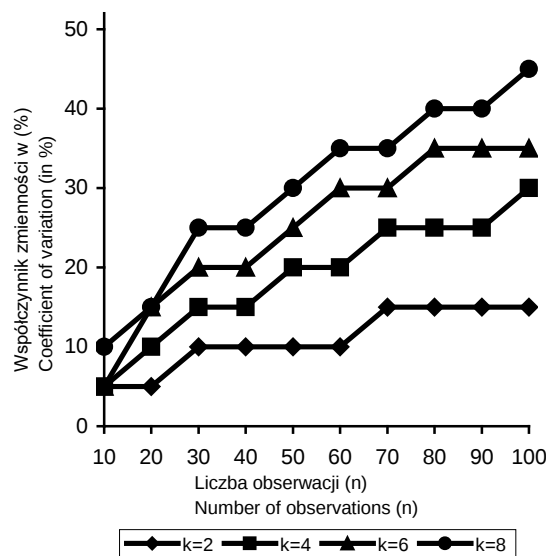
### Przykład 4

Testowanie na poziomie istotności  $\alpha = 0,05$  jednorodności czterech współczynników zmienności ( $H_0 : CV_1 = CV_2 = CV_3 = CV_4$ ) różniących się o 15% przy liczebnościach prób wynoszących 20. Mamy zatem:

$k = 4$ ;  $n_1 = n_2 = n_3 = n_4 = 20$ ;  $d = 15\%$ ;  $r = 5\%$ ;  $\alpha = 0,05$ .

Krok 1.  $CV_1 = 1\%$ ,  $CV_2 = 16\%$ ,  $CV_3 = 31\%$ ,  $CV_4 = 46\%$ ,  $Z = 54,023^*$  —  $H_0$  odrzucona  
 Krok 2.  $CV_1 = 6\%$ ,  $CV_2 = 21\%$ ,  $CV_3 = 36\%$ ,  $CV_4 = 51\%$ ,  $Z = 24,613^*$  —  $H_0$  odrzucona  
 Krok 3.  $CV_1 = 11\%$ ,  $CV_2 = 26\%$ ,  $CV_3 = 41\%$ ,  $CV_4 = 56\%$ ,  $Z = 15,766^*$  —  $H_0$  odrzucona  
 Krok 4.  $CV_1 = 16\%$ ,  $CV_2 = 31\%$ ,  $CV_3 = 46\%$ ,  $CV_4 = 61\%$ ,  $Z = 10,720^*$  —  $H_0$  odrzucona  
 Krok 5.  $CV_1 = 21\%$ ,  $CV_2 = 36\%$ ,  $CV_3 = 51\%$ ,  $CV_4 = 66\%$ ,  $Z = 7,706$  —  $H_0$  nieodrzucona  
 $\chi^2_{0,05; 2} = 7,815$ .

Wyniki badań symulacyjnych można także przedstawić graficznie wyznaczając dla różnych zbiorów współczynników zmienności pewne obszary krytyczne, tzn. obszary, w których zbiory  $k$  współczynników zmienności nie są jednorodne. Górne granice tych obszarów są wyznaczone dla kolejnych  $n_i = 10, 20, \dots, 100$  poprzez testowanie hipotezy o jednorodności  $k$  współczynników zmienności, z krokiem  $r = 5\%$ , poczynając od  $CV_1 = 1\%$  do momentu gdy wartość statystyki  $Z$  nie przekroczy wartości krytycznej  $\chi^2_{0,05; k-1}$ . Wówczas wartość najniższego współczynnika zmienności w zbiorze  $k$  współczynników jednorodnych określa górną granicę obszaru dla danego  $k$ . Rysunek 1 przedstawia „obszary krytyczne niejednorodności”  $k (= 2, 4, 6 \text{ i } 8)$  współczynników zmienności (różniących się o stałą wielkość  $d = 5\%$ ), wyznaczone przy poziomie istotności  $\alpha = 0,05$ .



Rys. 1. Obszary krytyczne o niejednorodności  $k$  ( $= 2, 4, 6, 8$ ) współczynników zmienności (różniących się o stałą wielkość  $d = 5\%$ ), wyznaczone przy poziomie istotności  $0,05$

Fig. 1. Critical areas of unhomogeneity of  $k$  ( $= 2, 4, 6, 8$ ) coefficients of variation (with the difference  $d = 5\%$ ), obtained at the significance level  $0.05$

#### WNIOSKI

1. Przy zachowaniu stałych liczb obserwacji  $n_i$ , stałej różnicy  $d$  między sąsiadującymi współczynnikami zmienności i stałej liczbie  $k$  współczynników zmienności, prawdopodobieństwo, że stanowią one będą zbiór jednorodny wzrasta wraz ze wzrostem wartości porównywanych współczynników.
2. Przy stałych liczbach obserwacji  $n_i$ , stałej różnicy  $d$  oraz takich samych minimalnych wartościach współczynników zmienności w danym zbiorze, prawdopodobieństwo, że zbiór  $k$  współczynników zmienności będzie jednorodny maleje wraz ze wzrostem liczby  $k$ .
3. Przy stałej liczbie  $k$  populacji, stałej różnicy  $d$  i ustalonej wartości najmniejszego w zbiorze współczynnika zmienności, prawdopodobieństwo, że zbiór będzie jednorodny maleje wraz ze wzrostem liczb obserwacji  $n_i$ .

#### LITERATURA

- Bennett B. M. 1976. On an approximate test for homogeneity of coefficients of variation., In: Contributions to applied statistics (ed. W. I. Ziegler). Birkhäuser Verlag: 169 — 171.
- Elandt R. 1964. Statystyka matematyczna w zastosowaniu do doświadczeń rolniczych. PWN, Warszawa.
- Greń J. 1987. Statystyka matematyczna. PWN. Warszawa.